Integration Metrics for Cartographic Generalization: Assessment of 1:1,000,000 Scale Hydrography and Terrain

Lawrence V. Stanislawski*, Barbara P. Buttenfield, Cynthia A. Brewer***, and Christopher Anderson-Tarver****

* Center of Excellence for Geospatial Information Science, U.S. Geological Survey, 1400 Independence Road Rolla MO 65401, USA
** Department of Geography, University of Colorado-Boulder, UCB-260 Boulder CO, 80309, USA
*** Department of Geography, Pennsylvania State University, University Park PA 16802, USA

Abstract. Cartographic generalization applies a series of operations to reduce map content and detail in a manner that legibly portrays desired features and conditions, often at a reduced scale. A variety of operations have been employed to maintain legibility and logical integration among feature themes. Techniques to evaluate generalization results can help automate generalization processing and the tuning of procedures and parameters. This paper presents methods to metrically evaluate how well a small scale hydrographic stream network conforms to an associated elevation dataset. For a sample of adjacent watersheds in the central United States, elevation-derived channels are extracted from the National Atlas terrain model at a density similar to the 1:1,000,000-scale (1M) National Atlas stream network. Horizontal displacement, content, and density distribution of the 1M linear stream network are metrically compared to the elevation-derived channels. Results from this research will help tune generalization processing for hydrography and terrain and develop better integration between these themes of the National Atlas and *The National Map*.

Keywords: cartographic generalization, integration, hydrography, terrain, national mapping

1. Introduction

Cartographic generalization reduces map content and detail in a manner that legibly portrays desired features and conditions, often at a reduced scale. Contextual generalization considers objects within their environment of surrounding features from possibly multiple themes, as compared to independent generalization that only involves an individual object or features within a single theme (Lecordix and Lemarie 2007). A variety of operations have been employed to maintain legibility and logical integration among feature themes. When generalizing roads and buildings to a reduced map scale for instance, graphic conflicts caused by enlarged symbols may be resolved through amalgamation, displacement, typification or other operations (Ware and Jones 1998, Regnauld 2001). Likewise, generalized hydrography or terrain may be re-shaped to create better visual integration between these two layers (Gaffuri 2007), and other techniques have been implemented for contextual generalization (Ruas and Duchene 2007, Lecordix and Lemarie 2007). Automated techniques to evaluate generalization results or to identify graphic conflicts have been used to automate generalization or to tune procedures and parameters (Regnauld 2001, Gaffuri 2007, Lecordix and Lemarie 2007, Ruas and Duchene 2007). This paper presents methods to metrically evaluate how well a stream network generalized to small scale conforms to an associated raster elevation dataset.

The United States Geological Survey (USGS) recently (2009) generalized the 1:100,000-scale (100K) National Hydrography Dataset (NHD) to form a 1:1,000,000-scale (1M) hydrography layer for the United States National Atlas® (Gary et al. 2010). This paper evaluates how well the linear stream network from the 1M National Atlas hydrography conforms to the associated National Atlas 100-meter resolution terrain data. For a sample of adjacent watersheds in central United States, elevation-derived channels are extracted from the National Atlas terrain model at a density similar to the 1M stream network. Horizontal displacement, content, and density distribution of the 1M linear stream network are metrically compared to the elevation-derived channels using the Coefficient of Linear Correspondence (CLC) and density-difference grids (Stanislawski et al. 2012). The 1M and elevation-derived stream networks are also compared with a more accurate stream network generalized from the high-resolution NHD (Buttenfield et al. 2013) to discern if either theme requires adjustment. It is expected that results of this research will help tune generalization processing and develop better integration among National Atlas themes, along with better integration between the National Atlas and *National Map* data.

2. Test Data and Methods

A set of 204 NHD subbasins that span a range of physiographic conditions around the Midwest region of the coterminous United States are used for this study (*Figure 1*). The 1M National Atlas stream lines within these subbasins are compared to drainage channels derived from the 100-meter (m) resolution digital elevation model (DEM) compiled for the National Atlas from the higher-resolution, 1 arc-second National Elevation Dataset (NED). The 1 arc-second NED has about 30 m cell resolution with 32-bit floating point values. It was converted to 16-bit integer values and projected from geographic coordinates to an Albers Equal-Area Conic projection with 100-m cells in the North American Datum of 1983. The result was clipped and extended to coastlines to form the 100-m National Atlas DEM, which is intended to portray terrain at 1M. All geoprocessing was completed with Esri ArcGIS® software.

Elevation-derived drainage channels are extracted from the 100-m National Atlas DEM using the weighted flow accumulation model applied by Stanislawski et al. (2012). The 1M National Atlas stream lines and the elevation-derived channels are compared through raster density differencing at 300-m resolution. Also, a conflation analysis was performed to compare the length of stream channels which match between the two datasets. Conflation results based on 1M buffering are summarized through the CLC metric.

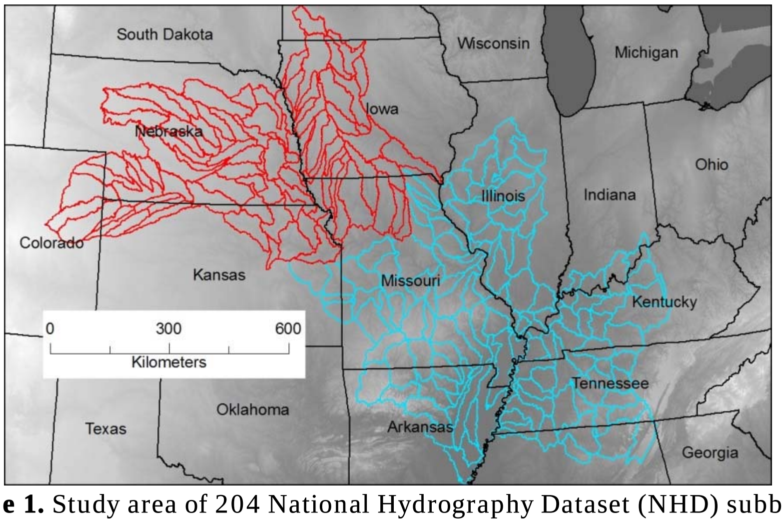


Figure 1. Study area of 204 National Hydrography Dataset (NHD) subbasins (outlined in red and light blue) in the middle of the United States. NHD Flowlines in northwest 83 subbasins (outlined in red) were generalized to 1:1,000,000-scale.

In addition, to approximate positional accuracy, a set of 1M level-of-detail (LoD) hydrographic network lines were compiled from the high-resolution (HR) NHD data for the northwest 83 subbasins in the study area (*Figure 1*). In the coterminous United States, the HR NHD is compiled from 1:24,000 or larger scale source data and likely has better positional accuracy than either the 1M National Atlas hydrography or the elevation-derived channels. Generalization of the HR NHD for these 83 subbasins follows procedures similar to those used to compile the 1M National Atlas hydrography (Buttenfield et al. 2013). The elevation-derived channels and the National Atlas hydrography lines are both compared to the 1M LoD hydrography flowlines by finding the proportion of the National Atlas and elevation-derived lines that fall within a series of different sized buffers around the LoD lines.

3. Results and Discussion

In the 204 subbasin study area, the 1M National Atlas includes about 117,410 kilometers (km) of hydrographic line features of type aqueduct, artificial path through polygonal water feature, canal, and intermittent and perennial stream (*Figure 2*). About 82,580 km of drainage channels were derived from the 100-m National Atlas elevation model for the study area, which represents about 40 percent less hydrographic lines than the National Atlas lines. The CLC conflation metric indicates that 65 percent of the National Atlas lines match the elevation-derived lines. A visual comparison indicates that most of the intermittent stream and canal features in the Atlas data are not represented in the elevation-derived drainage channels. Removing these features from the National Atlas features reduces the length of hydrographic lines to about 92,190 km, which is only about 10 percent longer than the length of the elevation-derived channels.

A more detailed visual inspection of the National Atlas hydrography lines (without intermittent streams and canals) and the elevation-derived lines (*Figure 3*) exposes the primary difference between the content of the two datasets, and the reason for miss-matches. The National Atlas lines (whitish lines) are intended to represent primary streams that are named or previously mapped at 1M, up to headwaters at the 100K representation (Gary et al. 2010). On the other hand, the elevation-derived channels (reddish lines) are more consistently mapped to the terrain based on accumulated drainage area, without bias for names or previously defined features. Additional channels exist in the elevation-derived lines that do not exist in the National Atlas lines. Where both sets of lines exist, the elevation-derived lines, in most cases, are not extended as far upstream as National Atlas lines.

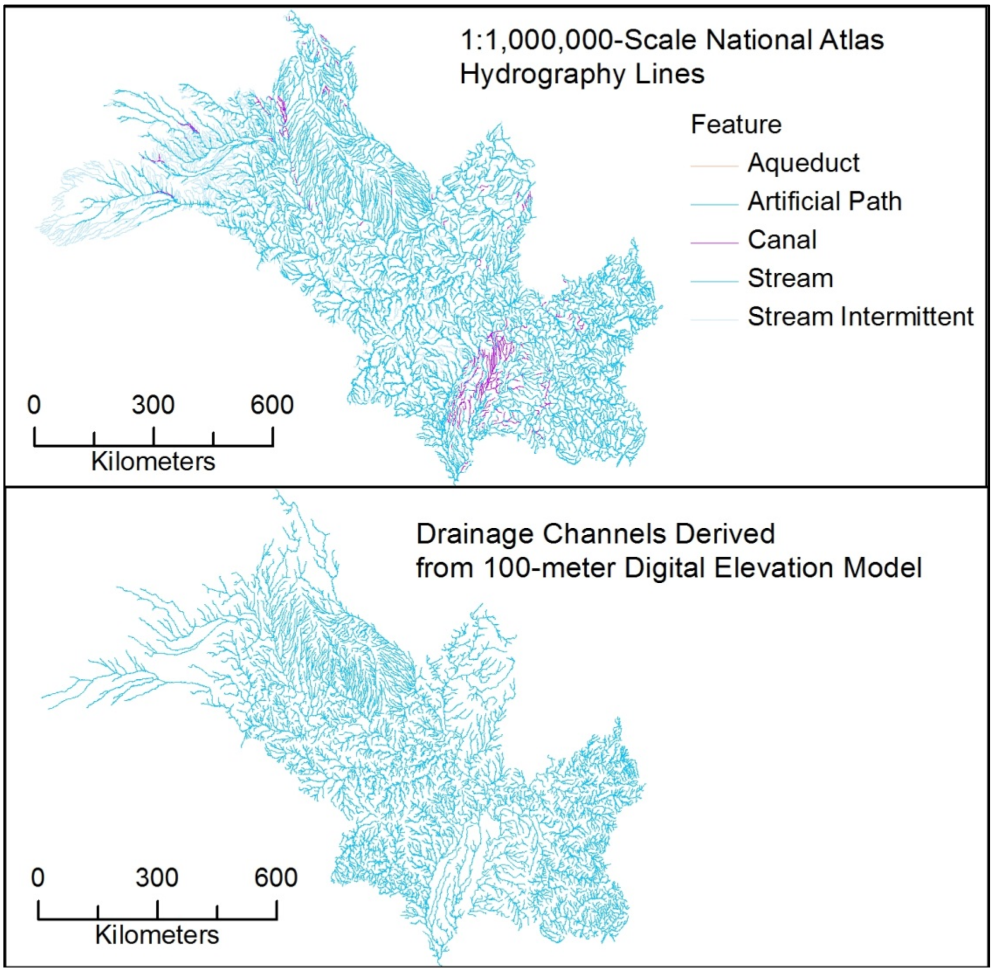
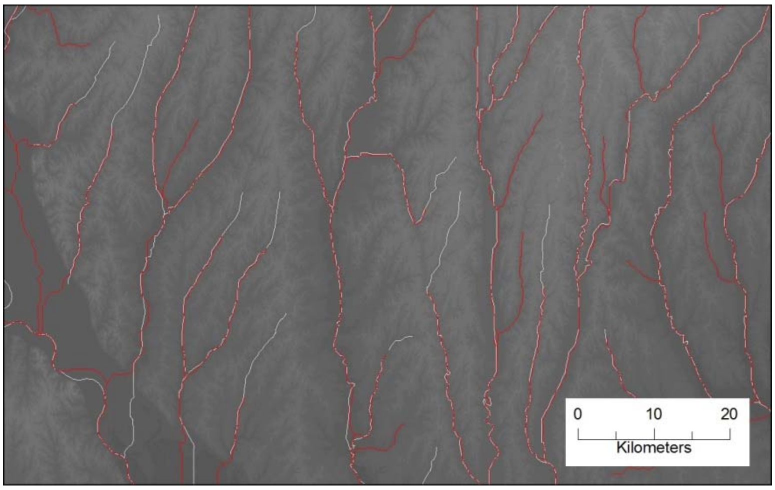


Figure 2. Feature types and distribution of 1:1,000,000-scale National Atlas hydrography lines (top), and drainage channels derived from 100-meter resolution elevation model (bottom) for the 204 subbasin area.

Another comparative analysis involved differencing raster density grids computed for the National Atlas line features and for the derived streams (*Figure 4*). The comparison was performed at 300-m resolution. The differencing reveals the spatial distributions where elevation-derived lines are omitted from the National Atlas lines (large negative density differences with reddish shade) and where National Atlas lines are omitted from the elevation-derived dataset (large positive differences with blue shade). The average value of the density difference grid that includes intermittent streams and canals in the Atlas density grid is 0.046 km per square km (km/km^2). When intermittent streams and canals are removed, the av-

erage density difference improves by more than 70 percent to only 0.013 km/km², indicating the content is more consistent between the two datasets.

Figure 3. National Atlas 1:1,000,000-scale (1M) hydrography lines (lighter whitish lines) and elevation-derived channels (darker reddish lines) displayed over 100-meter resolution elevation model at 1M. Intermittent streams and canals are re-



moved from National Atlas lines. Lines coincident between the two sets of lines appear orange.

Figure 5 graphs the proportion of the elevation-derived channels and National Altas lines that fall within various sized buffers around the (alleged more positionally accurate) 1M LoD lines that were generalized from the HR NHD. Results report the proportion of the target lines (i.e., National Atlas line and elevation-derived lines) that are within the buffers. Results do not consider the length of benchmark features (1M LoD) omitted from either set of target lines. CLC computations better reflect overall conflation between the two sets of lines. *Figure 5* indicates that nearly 90 percent or more of both sets of lines lie within buffers with radii of about 500 m or larger. This does **not** imply that both sets of lines are positionally within National Map Accuracy Standards (NMAS) of the 1M LoD lines. NMAS tolerance is 1/50th inch at map scale, which at 1M map scale, is about 508 m on the ground (US Bureau of the Budget 1947). The previous qualification is made for two reasons. First, in this assessement process, buffers around braided sections of channels, or relatively dense channel areas, may cause some false positive matches. Second, results are elevated because

downstream sections of unmatched channels confluing with matched channels fall within buffers and these sections are included as false-positives. Additional effort could refine procedures to exclude these false-positive situations.

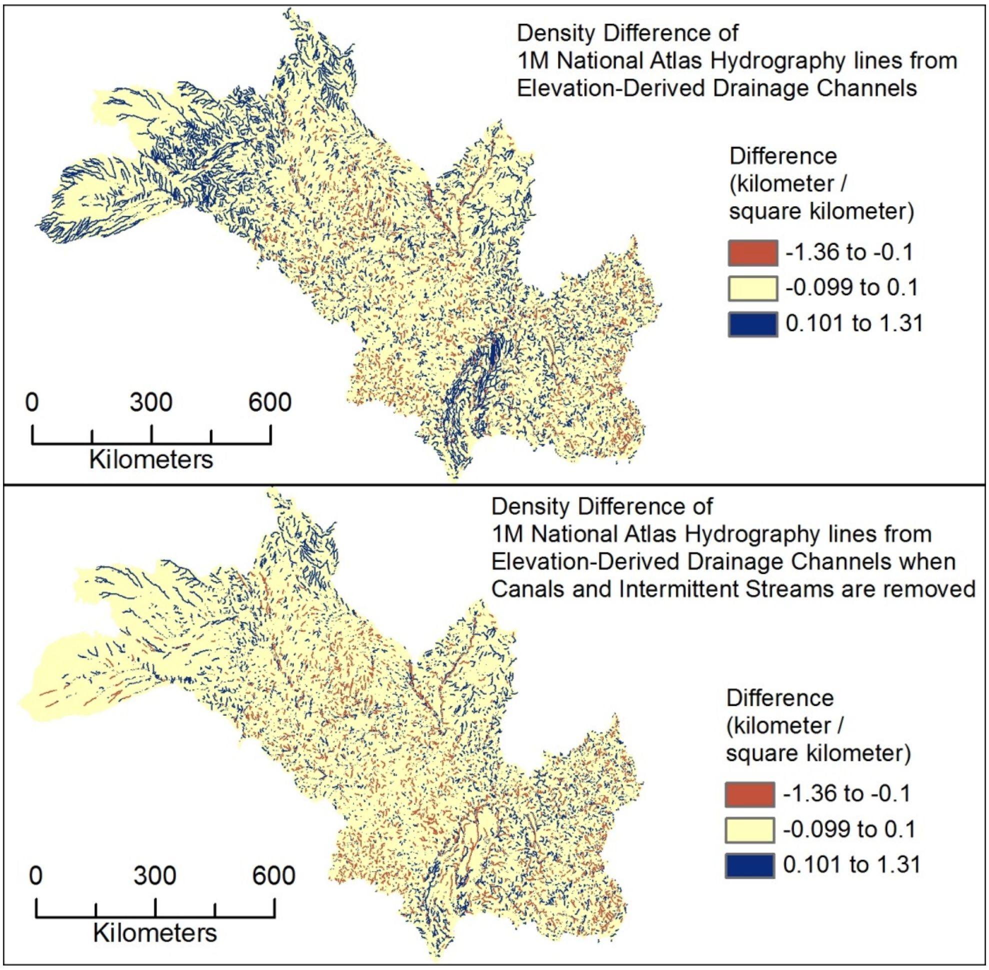


Figure 4. 300-meter resolution density difference raster datasets of the elevation-derived drainage channel density subtracted from the 1:1,000,000-scale (1M) National Atlas hydrography line density. Intermittent streams and canals are included in the National Atlas hydrography density in the top panel but removed for the density difference computed for the bottom panel.

Nevertheless, *Figure 5* presents an estimate of the amount of positional displacement for the National Atlas lines and the elevation-derived lines

relative to the 1M LoD lines. A more precise automated method is needed to quantify displacement between two sets of hydrographic lines. A final observation relates to *Figure 5*. The National Atlas lines retain more features positionally aligned with the 1M LoD lines than do the elevation derived lines for buffers less than 500 m. Considering that the generalization procedures for the 1M LoD lines were intended to mimic the National Atlas 1M lines, this result is expected and suggests a level of success in the 1M LoD generalization process.

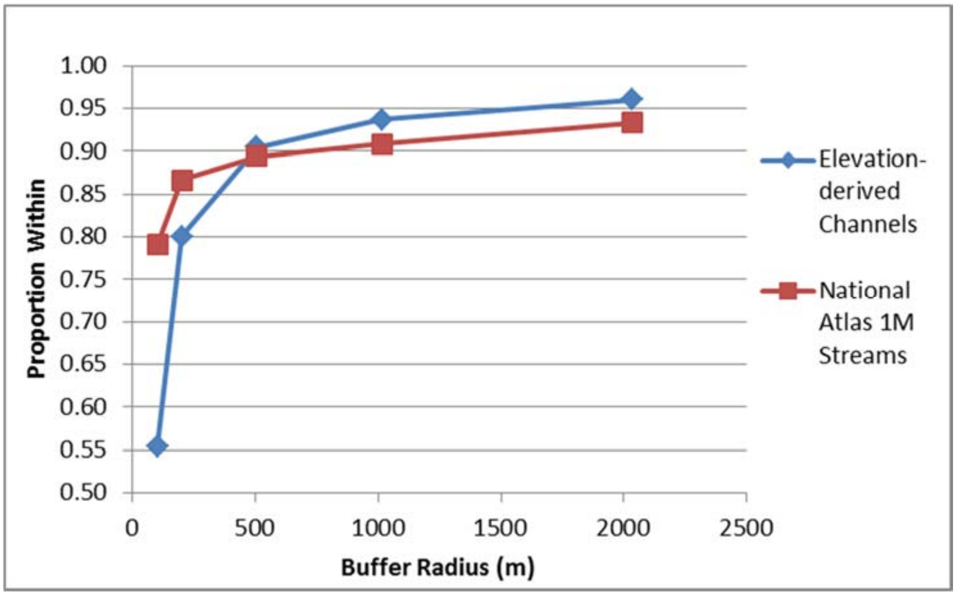


Figure 5. Proportion of elevation-derived drainage channels and National Atlas 1M hydrography lines that fall within buffers of various widths around the 1:1,000,000-scale level-of-detail (LoD) National Hydrography Dataset (NHD) flowlines.

Acknowledgements

Research funding provided by the U.S. Geological Survey's Center of Excellence for Geospatial Information Science (USGS-CEGIS) through the Department of Interior Cooperative Ecosystem Studies Unit (CESU). The work of Dr. Buttenfield and Chris Anderson-Tarver is supported by USGS-CEGIS grant #04121HS029, "Generalization and Data Modeling for New Generation Topographic Mapping." The work of Dr. Brewer is funded by grant #06HQAG0131, "Symbol and Label Design Solutions for Electronic Topographic Mapping for The National Map of the United States." We appreciate the support of Dr. E. Lynn Usery, Director of CEGIS.

References

Buttenfield B P, Stanislawski L V, Anderson-Tarver C, Gleason MJ (2013 submitted) Automatic Enrichment of Hydrographic Stream Networks with Primary Paths for the United States National Atlas. 26th International Cartographic Conference, Dresden, Germany

Buttenfield BP, Stanislawski LV, Brewer CA (2011) Adapting Generalization Tools to Physiographic Diversity for the United States National Hydrography Dataset. Cartography and Geographic Information Systems, 38(3):289–301

Gaffuri J (2007) Outflow Preservation of the Hydrographic Network on the Relief in Map Generalisation. 23rd International Cartographic Conference, Moscow, Russia

Gary RH, Wilson ZD, Archuleta C-AM, Thompson FE, Vrabel J (2010) Production of a National 1:1,000,000-Scale Hydrography Dataset for the United States—Feature Selection, Simplification, and Refinement. Scientific Investigation Report 2009-5202, U.S. Department of Interior, U.S. Geological Survey

Lecordix F, Lemarié C (2007) Managing Generalisation Updates in IGN Map Production. In W.A. Mackaness, A. Ruas, L.T. Sarjakoski (eds.), Generalization of Geographic Information: Cartographic Modeling and Applications, Elsevier for International Cartographic Association, 285-300

Regnauld N (2001) Contextual Building Typification in Automated Map Generalization, Algorithmica, 30: 312-333

Ruas A, Duchêne C (2007) A Prototype Generalisation System Based on the Multi-Agent System Paridigm. In WA Mackaness, A Ruas, LT Sarjakoski (eds.), Generalization of Geographic Information: Cartographic Modeling and Applications, Elsevier, 269-284

Stanislawski LV, Doumbouya AT, Miller-Corbett CD, Buttenfield BP, Arundel ST (2012) Scaling Stream Densities for Hydrologic Generalization. 7th International Conference on Geographic Information Science, Columbus, Ohio

Ware J M, Jones B J (1998) Conflict Reduction in Map Generalization Using Iterative Improvement, GeoInformatica 2(4):383-407

US Bureau of the Budget (1947). United States National Map Accuracy Standards, June 17.